ORIGINAL ARTICLE

Examining weekly heart rate variability changes: a comparison between monitoring methods

Santiago A. Ruiz-Alias[1,2] · Aitor Marcos-Blanco[1,2] · Antonio Clavero-Jimeno[1,2] · Felipe García-Pinillos[1,2,3]

Abstract

Monitoring heart rate variability has been commonly performed by different devices which differ in their methods (i.e., night recording vs. upon awakening measure, pulse vs. R waves, and software signal processing), Thus, the purpose of this study was to determine the level of agreement between different methods of heart rate variability monitoring, represented in two different systems (i.e., the Polar Nightly Recharge™ function present in Polar sport watches and the Polar H10 chest strap synchronized with the Kubios app). A group of 11 recreational athletes performed a concurrent training program for eight weeks and heart rate variability was daily monitored through both devices. Very large correlation ($r = 0.714$) and good reliability (ICC = 0.817) were obtained between devices through the entire training program. The magnitude-based inference method was also applied to determine the likelihood of the change concerning the smallest worthwhile change. From a baseline corresponding to the first two weeks of the training program, the weekly heart rate variability changes of the following six weeks were determined for each participant with each device. Despite the large correlation and good reliability between devices, there was a 60.6% of discordance in the likelihood interpretation of the change for the 66 weeks evaluated, explained by the random errors found. Thus, practitioners should be aware of these differences if their training groups use different devices or if an athlete interchanges them. The different nuances of each device can condition the heart rate variability data variation which could compromise the interpretation of the autonomic nervous system modulation.

Abbreviations

HRV	Heart rate variability
ANS	Autonomic nervous system
SNS	Sympathetic nervous system
PNS	Parasympathetic nervous system
SWS	Slow-wave sleep
PPG	Photoplethysmography
PRV	Pulse rate variability
SWC	Smallest worthwhile change
MBI	Magnitude-based inference

✉ Santiago A. Ruiz-Alias
  aljruiz@ugr.es

1 Department of Physical Education and Sport, University of Granada, Carretera de Alfacar s/n, 18011 Granada, Spain

2 Sport and Health University Research Center (iMUDS), C/ Menéndez Pelayo 32, 18016 Granada, Spain

3 Department of Physical Education, Sports and Recreation, Universidad de La Frontera, Temuco, Chile

1 Introduction

Athletes' monitoring has been one of the key aspects of the training process to control the balance between external and internal loads. In this manner, professionals in the field can maximize performance by reducing the potential risks of illness and injury [1]. The continuous innovations in monitoring technologies have created a wide variety of tools that are accessible nowadays to everyone. For this purpose, heart rate variability (HRV) has been one of the most used metrics due to its non-invasive, inexpensive, and time-efficiency [1]. HRV reflects the autonomic nervous system

(ANS) modulation by the variation of the inter-beat intervals commonly expressed as time-domain (RMSSD, SDNN, PNN50%), frequency-domain (VLF, LF, HF), and nonlinear variables (SD1, SD2) [2]. Each of them corresponds to a branch of the ANS, either the sympathetic (SNS) or the parasympathetic activity (PNS) [3]. From all of these options, Buccheit [3] and Plews et al. [4] recommended practitioners to use the natural logarithm of the time domain called the square root of the mean of the sum of the squares of differences between adjacent normal R-R intervals (Ln RMSSD) for a couple of reasons: (i) it can be captured by a short record, (ii) and when compared the frequency domain, the Ln RMSSD is less sensitive to breathing noises what makes it ideal for daily monitoring without requiring a controlled breathing pattern. From a baseline, changes in Ln RMSSD can be interpreted mainly as coping well with training when the activity of the PNS increases, and as accumulated fatigue when Ln RMSSD decreases as a reflection of the increased SNS activity [3].

Some methodological and practical considerations have been remarked as crucial to get valid and reliable HRV data in a monitoring context [3]. Given the high sensitivity of the ANS to environmental conditions, resting HRV measurements must be conducted under standardized conditions [3]. Thus, nightly records during the first hours of sleep, where the third phase called the slow-wave sleep (SWS) mostly occurs, constitute a theoretically ideal moment for HRV monitoring [3]. However, a late-night exercise on the previous day and a stage sleeping fragmentation can alter the Ln RMSSD value despite perceiving a well rest [3]. For this reason, upon awakening HRV measurement can be a more appropriate method if a standardized environment is applicated such as same bed, same hour, and no daily activity done before measure [3].

A wide variety of sport watches and mobile phones apps have offered this HRV tracking function. One of the most used by the sports community is the Polar Nightly Recharge™ function present in Polar sport watches [5]. By the technique of photoplethysmography (PPG), the ANS activity can be monitored by the pulse rate variability (PRV) [6] during the first four hours of sleep [5]. On the other hand, electrocardiogram R waves have been the principal source for measuring HRV (i.e., gold-standard measurement) and it can be easily tracked nowadays in a valid way by a chest strap [7]. Another crucial aspect to consider is the signal processing since any ectopic beat can alter significantly the HRV value [3]. The Kubios software available in a mobile phone app has shown to be a valid tool in this aspect by interpolating the identified artefacts using a cubic spline interpolation [8]. Thus, the combination of a chest strap with the app Kubios is also a common practice in the sport community to track HRV upon awakening.

Since both methods are widely used by athletes and practitioners, it is necessary to compare both systems to clarify their potential differences and similarities. Due to the aforementioned disparities between methods (i.e., night vs. upon awakening measure, pulse vs. R waves, and Polar vs. Kubios software signal processing) and the individuality of HRV values, the most appropriate way to determine their level of agreement is to analyze the individual response of each athlete along a training program with both devices.

To this end, the weekly Ln RMSSD fluctuation represents more consistently than daily values any change of the cardiac ANS modulation [9], and by the method proposed by Batterham and Hopkins [10] called magnitude-based inference (MBI), the ANS modulation can be determined by the smallest worthwhile change (SWC) at each device. This procedure consists of determining a baseline HRV value by tracking a light training period and then comparing it to the following microcycles of the training program. Based on the magnitude and variability of the change with respect to the baseline value where the SWC comes from, Batterham and Hopkins [10] proposed to interpret any change, in this case, an increase or decrease of the PNS activity as indicated in Table 1 to determine how the athlete is coping with training. This novel approach would allow a comparison between both devices despite their different tracking procedures and determine their possible similarities based on their interpretations of the change.

Based on the aforementioned, the purpose of this study was to determine the level of agreement between different methods of HRV monitoring, represented in two different systems (i.e., the Polar Nightly Recharge™ function present in Polar sport watches and the Polar H10 chest strap synchronized with the Kubios app) in the weekly Ln RMSSD changes during an eight-week concurrent training program in recreational athletes. Based on the different methodological considerations (i.e., night recording vs. upon awakening measure, pulse vs. R waves, and software signal processing), we hypothesized that HRV data obtained from both devices would be correlated but they would differ in absolute values which would influence the interpretation of the weekly Ln RMSSD changes.

Table 1 Magnitude-based inference method change interpretation [10]

Likelihood	Interpretation
<0.5%	Most unlikely
0.5–5%	Very unlikely
5–25%	Unlikely
25–75%	Possibly
75–95%	Likely
95–99.5%	Very likely
>99.5%	Most likely

2 Methods

All participants were informed about the research purpose and procedures of the study before signing a written informed consent form. The study protocol adhered to the tenets of the Declaration of Helsinki and was approved by the institutional review board (ref. 2546/CEIH). To address the aim of determining the level of agreement between different methods of HRV monitoring, the data collection, process, and comparison were executed in summary as follow:

Data collection: Participants' HRV (RMSSD) were monitored during an 8-week training program using two different systems (i.e., the Polar Nightly Recharge™ function present in Polar sport watches and the Polar H10 chest strap synchronized with the Kubios app).

Data processing: RMSSD data were transformed to Ln RMSSD and averaged over each week (Weekly Ln RMSSD) of the training program following the recommendation of Plews et al. [4, 9].

Data comparison: The relationship, absolute agreement, systematic bias, and random error between devices were determined through Pearson correlation, intra-class correlation coefficients (ICC), and Bland Altman plots for the 8 weeks dataset (Ln RMSSD). The weekly Ln RMSSD changes interpretation during the eight-week concurrent training program was also determined through the MBI in both devices and established in a cross table.

2.1 Study design

All participants completed an eight-week concurrent training program composed of three sessions of 60–90 min per week with all-out sprint intervals and squats and bench press strength exercises (Fig. 1). Individualized loads were applied according to a one-maximum repetition test for the aforementioned exercises. The training program progresses every two weeks in volume

for the all-out sprint intervals and in load for the strength exercises (Fig. 1). During the training program, athletes were asked to self-monitor their HRV with the Nightly Recharge™ function present in the Polar Ignite sport watch (Version number: 2.1.5) and the with the Polar H10 chest strap (Version number: 3.1.1) synchronized to the app Kubios (Version number: 1.1.10). The first two weeks of the training program were used to determine the SWC (mean ± 0.2SD) where the lightest micro-cycle load was applied, to determine in this way, the weekly Ln RMSSD changes on the following six weeks of the training program (Fig. 1).

2.2 Participants

A group of 18 recreational athletes was initially enrolled. After checking the compliance with the intervention period, seven of them were discarded as they did not conduct at least three measures per week as Plews et al. [11] recommended to have a valid representation of one-week daily monitoring. Finally, a group of 11 recreational athletes, five females and six males (age: 21.73 ± 1.49 years, body mass: 70.74 ± 12.88 kg, height: 173 ± 7.6 cm, body mass index: 23.41 ± 2.99 kg/m^2) were involved in their weekly HRV changes analysis. A post hoc analysis of the achieved power for this sample size was conducted (G*Power software vs. 3.1), given $\alpha = 0.05$, $r = 0.7$, total sample size $= 11$, statistical test $=$ Correlation bivariate normal model. This analysis revealed a medium to large power (0.7). All participants met the following inclusion criteria: (i) to not suffer any injuries within the six months before data collection, (ii) to be physically active according to the guidelines of the ACSM [12].

2.3 Data collection and processing

The HRV was self-monitored by athletes with the Nightly Recharge™ function present in the Polar Ignite sport watch and with the Polar H10 chest strap synchronized
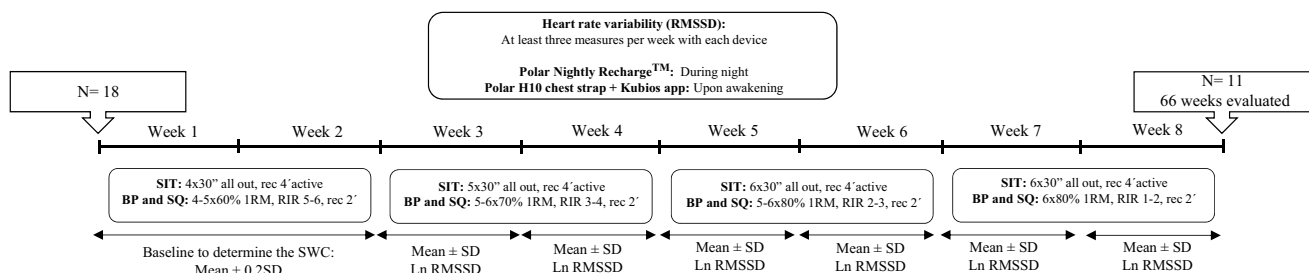

Fig. 1 Training program diagram followed to test the level of agreement in the HRV monitoring between the Polar H10 chest strap synchronized with the Kubios app and the Polar Nightly Recharge™ function present in the Polar Ignite sport watch. *SWC* Smallest worthwhile change, *SIT* Sprint interval training, *BP* Bench press, *SQ* Squat,

1RM One-repetition maximum, *RIR* Repetitions in reserve, *RMSSD* Root of the mean of the sum of the squares of differences between adjacent normal R–R intervals, *Ln RMSSD* natural logarithm of RMSSD, *SD* standard deviation, *Rec* recovery

to the app Kubios. They were asked to place the Polar Ignite sport watch on their non-dominant arm wrist and to adjust the strap until a tight feeling to avoid any motion artifacts [13]. Athletes were also instructed to leave the chest strap on the bedside to avoid any disturbance to conduct the upon awakening measure with the Polar H10 chest strap synchronized to the app Kubios in a supine position with a short-term record of five minutes [14]. On an excel shared spreadsheet, participants registered their daily HRV data corresponding to the (RMSSD) value from both devices along the training program which was then transformed into (Ln RMSSD) and averaged for each week (Weekly Ln RMSSD) following the recommendations of Plews et al. [4, 9].

2.4 Statistical analysis

Descriptive statistics are represented as mean (SD). The normal distribution of data and homogeneity of variances were confirmed through the Shapiro–Wilk test and Levene's tests, respectively ($p > 0.05$). The relationship between the HRV data (Ln RMSSD) obtained through both devices was quantified through the Pearson correlation coefficient. The criteria for interpreting the magnitude of the coefficients were as follows: trivial (0.00–0.09), small (0.10–0.29), moderate (0.30–0.49), large (0.50–0.69), very large (0.70–0.89), nearly perfect (0.90–0.99), and perfect (1.00) [15]. The Bland–Altman method was used to determine the systematic bias and random error (mean difference $\pm$ 1.96 SD) [16]. ICC was also calculated according to the guidelines reported by Koo and Li [17]. A two-way mixed effects model was performed following a mean of two raters type (Kubios-H10 vs. Polar Ignite) to determine their absolute agreement. The 95% confidence intervals of the ICC values were calculated, and interpretations were based on the following classifications: < 0.50 (poor reliability), 0.50–0.75 (moderate reliability), 0.75–0.90 (good reliability), and > 0.90 (excellent reliability) [17]. The level of significance used was $p < 0.05$. Data analysis was performed using the SPSS software (version 21, SPSS Inc., Chicago, IL, USA). The MBI method was used to determine the weekly HRV change interpretation of each participant with each device [10]. The data collected over the first two weeks of the training program with each device were used to establish the SWC of each participant (mean $\pm$ 0.2SD) [18]. Using Hopkin's spreadsheet (http://www.sportsci.org) [19], the likelihood of a true change was interpreted as indicated in Table 1 [10]. If there was a possibly decrease and increase, the interpretation was established as unclear [10]. Lastly, the magnitude-based weekly Ln RMSSD changes interpretation of each participant with each device was summarized on a cross table to determine the percentage of the agreement for the 66 weeks evaluated.

3 Results

A very large correlation ($r = 0.714$) and good reliability (ICC $= 0.817$) were obtained between devices through the entire training program (Table 2). A higher coefficient of variation was found in the Polar H10-Kubios (9.93%) compared to the Polar Ignite (7.71%) for the eight weeks training program. In Bland–Altman plots (Fig. 2), males and females presented a similar systematic bias and random error of -0.05 ± 0.33 ms and -0.01 ± 0.28 ms, respectively.

Relative to the individual HRV analysis, Table 3 summarizes the magnitude-based weekly Ln RMSSD changes of each participant with each device during 8-week training program. Based on the SWC obtained from the initial 2 weeks (mean $\pm$ 0.2SD), the area where the 90% of confidence interval overlaps determines the likelihood of the change (Fig. 3). In the example exposed, devices only agree

Table 2 Pearson (r), intra-class correlation coefficients (ICCs), and coefficient of variation (CV) of the heart rate variability (Ln RMSSD, in ms) data obtained from two different devices (i.e. Kubios-H10 vs. Polar Ignite) during an eight week concurrent training program

Variable	Pearson (r)	ICC (95% CI)	CV (%)	
			H10-Kubios	Ignite
Ln RMSSD (ms)				
Week 1	0.691***	0.817 (0.711–0.884)	8.09	7.56
Week 2	0.764***	0.841 (0.740–0.902)	9.86	7.41
Week 3	0.733***	0.837 (0.737–0.900)	10.78	8.94
Week 4	0.811***	0.887 (0.819–0.929)	9.17	7.74
Week 5	0.670***	0.791 (0.658–0.872)	10.19	7.99
Week 6	0.664***	0.755 (0.593–0.851)	10.22	7.27
Week 7	0.738***	0.803 (0.688–0.876)	11.24	7.21
Week 8	0.707***	0.812 (0.535–0.924)	10.08	7.36
♂♀Week 1–8	0.714***	0.817 (0.782–0.846)	9.93	7.71
♂ Week 1–8	0.730***	0.751 (0.684–0.804)	10.25	5.41
♀ Week 1–8	0.718***	0.836 (0.787–0.873)	8.54	8.81

***Indicates statistical significance for the Pearson correlation analysis ($p < 0.001$); Ln RMSSD: natural logarithm of the square root of the mean of the sum of the squares of differences between adjacent normal R–R intervals; ♂♀: males and females ($n = 11$); ♂: males ($n = 6$); ♀: females ($n = 5$)
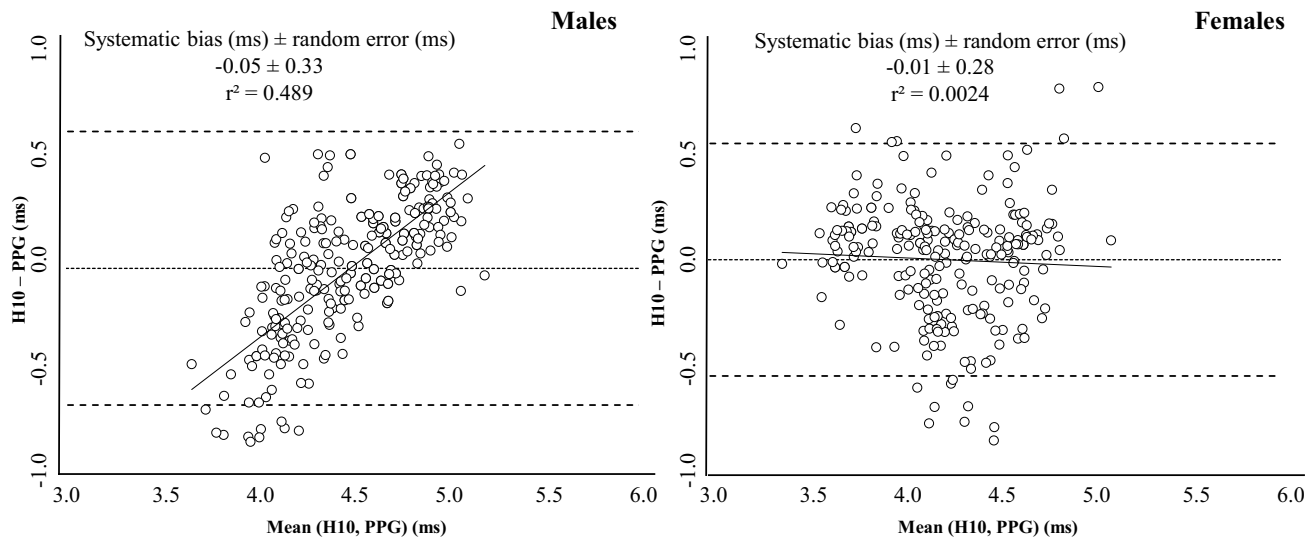
Fig. 2 Bland–Altman plots with the mean differences of heart rate variability (Ln RMSSD, in ms) obtained through the eight-week concurrent training program from two different devices (Kubios-H10 vs. Polar Ignite). The plots include the mean difference (dotted line), 95% limits of agreement (dashed lined), and the regression line (solid line). Systematic bias and Pearson's determination coefficient (r^2) are also presented

in week five in the direction of the change but with a discordance in the likelihood (Fig. 3).

Table 4 summarizes the magnitude-based weekly Ln RMSSD changes interpretation of each participant with each device. The white diagonal squares represent the agreement between devices for the likelihood and the direction of the change. The dark grey squares surrounding the white diagonal indicate the weeks where devices agree in direction but differ in the likelihood of the change. The absolute agreement between devices for the 66 weeks evaluated was 13.6%, meanwhile, the agreement in the direction of the change but with discordance in likelihood was 25.8%.

4 Discussion

Supporting our main hypothesis, the results reveal that despite the very large correlation ($r = 0.714$) and good reliability (ICC $= 0.817$) obtained between devices through the entire training program, the MBI method applied to determine the weekly Ln RMSSD changes from the 66 weeks monitored showed an absolute agreement of 13.6%, and an agreement in the direction of the change but with discordance in the likelihood of 25.8% (Table 3). These systems use different methods to get the HRV data and some methodological differences might explain the low level of agreement between devices when the MBI method was applied. Whereas the Polar Nightly Recharge™ records at the first four sleeping hours corresponding to SWS [5], the Polar H10 synchronized with the Kubios app is applied upon awakening. Regarding this, Hynynen et al. [20] found in a

group of 12 overtraining female and male athletes, that the ANS modulation was disturbed upon awakening measures but not during sleep with respect to the control group. A potential reason is that awakening itself involves a sympathetic stimulus, which could represent in a better way the athlete´s state than the sleeping record [20]. Another cause of disparity could be the sleep stage distribution as it can alter the Ln RMSSD value despite having a perceived well rest [3]. Since the Polar Nightly Recharge™ record at the first four sleeping hours, a greater percentage time of the third sleep stage (SWS) during them could involve greater Ln RMSSD values compared to other nights with the same amount of stage time but in different time frames [3]. On the other hand, the ANS modulation during these first sleeping hours can be also affected by the late-night exercise [21]. In our case, some athletes conducted their training sessions between 20:00 and 21:00 h which could have involved an alteration of the measure for those days.

Another important point comes from the data source. Polar devices include the PPG Polar Precision Prime™ system that detects the pulse waves of the cardiac output, meanwhile, their Polar H10 chest strap receives the R waves of the heartbeat electrical impulses. The main source of error of PPG devices are the motion artefacts that are mainly present in physical activities, meanwhile at rest, they present a reasonable accuracy [22]. Since these devices have been tested in different running intensities showing a high level of agreement [23, 24], this potential source of inaccuracy seems to be well controlled with just a correct strap adjustment [13]. The Polar H10 chest strap has shown to be a valid tool compared to the electrocardiogram for measuring the

Table 3 Magnitude-based weekly Ln RMSSD changes during an eight-week concurrent training program determined by two different devices (i.e. Kubios-H10 vs. Polar Ignite)

Subject		Week 1–2 (baseline) Mean±SWC	Week 3 Mean±SD	Week 3 % chances ↓/↔/↑	Week 4 Mean±SD	Week 4 % chances ↓/↔/↑	Week 5 Mean±SD	Week 5 % chances ↓/↔/↑	Week 6 Mean±SD	Week 6 % chances ↓/↔/↑	Week 7 Mean±SD	Week 7 % chances ↓/↔/↑	Week 8 Mean±SD	Week 8 % chances ↓/↔/↑
1	K	3.80±0.06	3.79±0.24	34/34/32 Unclear	3.78±0.34	44/18/37 Unclear	4.04±0.40	19/10/70 Possibly ↑	3.86±0.23	20/30/49 Possibly ↑	3.83±0.21	24/35/41 Possibly ↑	3.87±0.07	2/42/55 Possibly ↑
	I	3.99±0.04	3.87±0.35	64/12/24 Possibly ↓	4.14±0.15	3/9/88 Likely ↑	4.19±0.09	0/1/99 Very likely ↑	4.04±0.12	12/33/55 Possibly ↑	3.91±0.15	66/23/11 Possibly ↓	4.11±0.10	3/11/86 Likely ↑
2	K	4.13±0.02	3.98±0.01	100/0/0 Most Likely ↓	4.34±0.04	0/0/100 Most Likely ↑	4.26±0.10	2/3/95 Very likely ↑	4.29±0.09	0/1/98 Very likely ↑	4.32±0.07	0/0/100 Most likely ↑	4.23±0.05	0/3/97 Very likely ↑
	I	4.26±0.03	4.28±0.10	25/35/40 Unclear	4.35±0.06	1/5/94 Likely ↑	4.34±0.06	1/9/90 Likely ↑	4.35±0.08	3/10/88 Likely ↑	4.44±0.09	1/1/98 Very likely ↑	4.30±0.07	10/29/61 Possibly ↑
3	K	4.97±0.03	5.05±0.09	4/17/79 Likely ↑	5.04±0.14	13/20/67 Possibly ↑	4.99±0.11	22/36/42 Possibly ↑	5.05±0.06	1/8/91 Likely ↑	5.14±0.10	1/2/98 Very likely ↑	5.07±0.19	17/13/69 Likely ↑
	I	4.84±0.02	4.79±0.07	72/21/7 Possibly ↓	4.94±0.20	16/7/74 Possibly ↑	4.74±0.13	83/9/7 Likely ↓	4.80±0.07	68/24/8 Possibly ↓	4.81±0.07	59/31/10 Possibly ↓	4.80±0.07	69/22/9 Possibly ↓
4	K	4.79±0.02	4.65±0.18	85/7/9 Likely ↓	4.82±0.17	33/19/49 Unclear	4.65±0.14	89/6/6 Likely ↓	4.69±0.17	81/9/10 Likely ↓	4.64±0.18	89/5/5 Likely ↓	4.66±0.10	94/4/2 Likely ↓
	I	4.61±0.03	4.62±0.17	34/24/42 Unclear	4.66±0.12	14/27/59 Possibly ↑	4.57±0.05	68/30/2 Possibly ↓	4.67±0.12	12/25/63 Possibly ↑	4.64±0.07	9/46/45 Possibly ↔	4.59±0.16	49/22/29 Unclear
5	K	4.03±0.03	3.86±0.10	81/8/11 Likely ↓	4.11±0.11	25/13/62 Unclear	3.99±0.09	52/18/30 Possibly ↓	3.98±0.09	52/17/26 Unclear	3.69±0.09	97/1/1 Very likely ↓	3.73±0.17	84/4/11 Likely ↓
	I	4.27±0.02	4.27±0.03	34/33/34 Unclear	4.23±0.05	65/16/19 Possibly ↓	4.23±0.07	59/14/27 Unclear	4.28±0.07	41/15/44 Unclear	4.33±0.06	19/14/68 Possibly ↑	4.27±0.06	40/19/41 Unclear
6	K	4.28±0.04	4.26±0.03	34/50/17 Possibly ↔	4.12±0.07	87/9/5 Likely ↓	4.10±0.12	77/9/13 Likely ↓	4.05±0.08	93/4/3 Likely ↓	4.04±0.07	96/3/1 Very likely ↓	4.14±0.21	62/9/29 Unclear
	I	4.41±0.05	4.39±0.07	36/36/28 Unclear	4.38±0.05	40/41/19 Possibly ↔	4.31±0.05	74/21/5 Possibly ↓	4.39±0.04	27/58/15 Possibly ↔	4.30±0.05	79/18/4 Likely ↓	4.39±0.12	42/21/37 Unclear
7	K	4.14±0.08	4.18±0.17	31/24/45 Unclear	4.31±0.14	13/20/67 Possibly ↑	4.26±0.07	4/29/67 Possibly ↑	4.10±0.20	45/21/35 Unclear	4.13±0.08	29/47/24 Possibly ↔	4.25±0.06	3/31/66 Possibly ↑
	I	4.37±0.04	4.44±0.18	16/22/62 Possibly ↑	4.30±0.07	63/21/16 Possibly ↓	4.26±0.05	82/14/5 Likely ↓	4.33±0.10	51/20/30 Unclear	4.36±0.08	42/25/34 Unclear	4.38±0.07	32/27/41 Unclear

Table 3 (continued)

Subject		Week 1–2 (baseline) Mean±SWC	Week 3 Mean±SD	% chances ↓/↔/↑	Week 4 Mean±SD	% chances ↓/↔/↑	Week 5 Mean±SD	% chances ↓/↔/↑	Week 6 Mean±SD	% chances ↓/↔/↑	Week 7 Mean±SD	% chances ↓/↔/↑	Week 8 Mean±SD	% chances ↓/↔/↑
8	K	3.87±0.03	3.74±0.05	89/7/4 Likely ↓	3.74±0.08	78/11/11 Likely ↓	3.75±0.03	97/3/0 Very likely ↓	3.75±0.06	84/10/6 Likely ↓	3.87±0.06	37/24/39 Unclear	3.86±0.08	43/20/36 Unclear
	I	3.74±0.03	3.66±0.05	71/19/9 Possibly ↓	3.64±0.04	85/12/4 Likely ↓	3.66±0.02	95/5/0 Very likely ↓	3.69±0.03	67/28/4 Possibly ↓	3.74±0.05	35/34/30 Unclear	3.73±0.04	35/38/27 Unclear
9	K	4.25±0.03	4.31±0.09	25/15/60 Possibly ↑	4.20±0.06	60/20/20 Possibly ↓	4.30±0.03	7/25/67 Possibly ↑	4.29±0.05	18/24/58 Possibly ↑	4.27±0.12	39/13/48 Unclear	4.19±0.15	56/10/34 Unclear
	I	4.18±0.02	4.28±0.04	3/8/89 Likely ↑	4.22±0.03	11/28/61 Possibly ↑	4.33±0.04	1/2/97 Very likely ↑	4.17±0.05	41/26/33 Unclear	4.30±0.04	2/5/93 Likely ↑	4.24±0.06	16/16/68 Possibly ↑
10	K	4.08±0.01	4.07±0.03	50/25/25 Unclear	4.09±0.03	32/21/47 Unclear	4.18±0.03	2/3/95 Very likely ↑	4.03±0.02	87/9/3 Likely ↓	4.07±0.04	50/16/33 Unclear	4.16±0.02	1/3/96 Very likely ↑
	I	4.17±0.04	4.10±0.06	61/27/11 Possibly ↓	4.04±0.07	79/14/7 Likely ↓	3.96±0.07	92/5/3 Likely ↓	4.36±0.04	0/1/98 Very likely ↑	4.31±0.10	12/14/74 Possibly ↑	4.13±0.08	51/26/23 Possibly ↓
11	K	4.76±0.05	4.73±0.04	40/49/11 Possibly ↔	4.59±0.04	95/4/0 Very likely ↓	4.68±0.03	67/31/2 Possibly ↓	4.63±0.04	91/8/1 Likely ↓	4.67±0.02	86/14/0 Likely ↓	4.67±0.03	81/18/1 Likely ↓
	I	4.62±0.04	4.71±0.04	4/23/73 Possibly ↑	4.63±0.06	29/36/34 Unclear	4.55±0.02	81/18/0 Likely ↓	4.55±0.06	73/25/1 Possibly ↓	4.61±0.01	2/98/0 Very likely ↔	4.56±0.03	62/35/3 Possibly ↓

<0.5%: most unlikely; 0.5–5%: very unlikely; 5–25%: unlikely; possibly: 25–75%; likely: 75–95%; very likely: 95–99.5%; most likely: >99.5%; unclear: possibly ↓/↑ ↓: decrease; ↑: increase; ↔: unchanged; SWC: mean±0.2SD (baseline week 1–2)
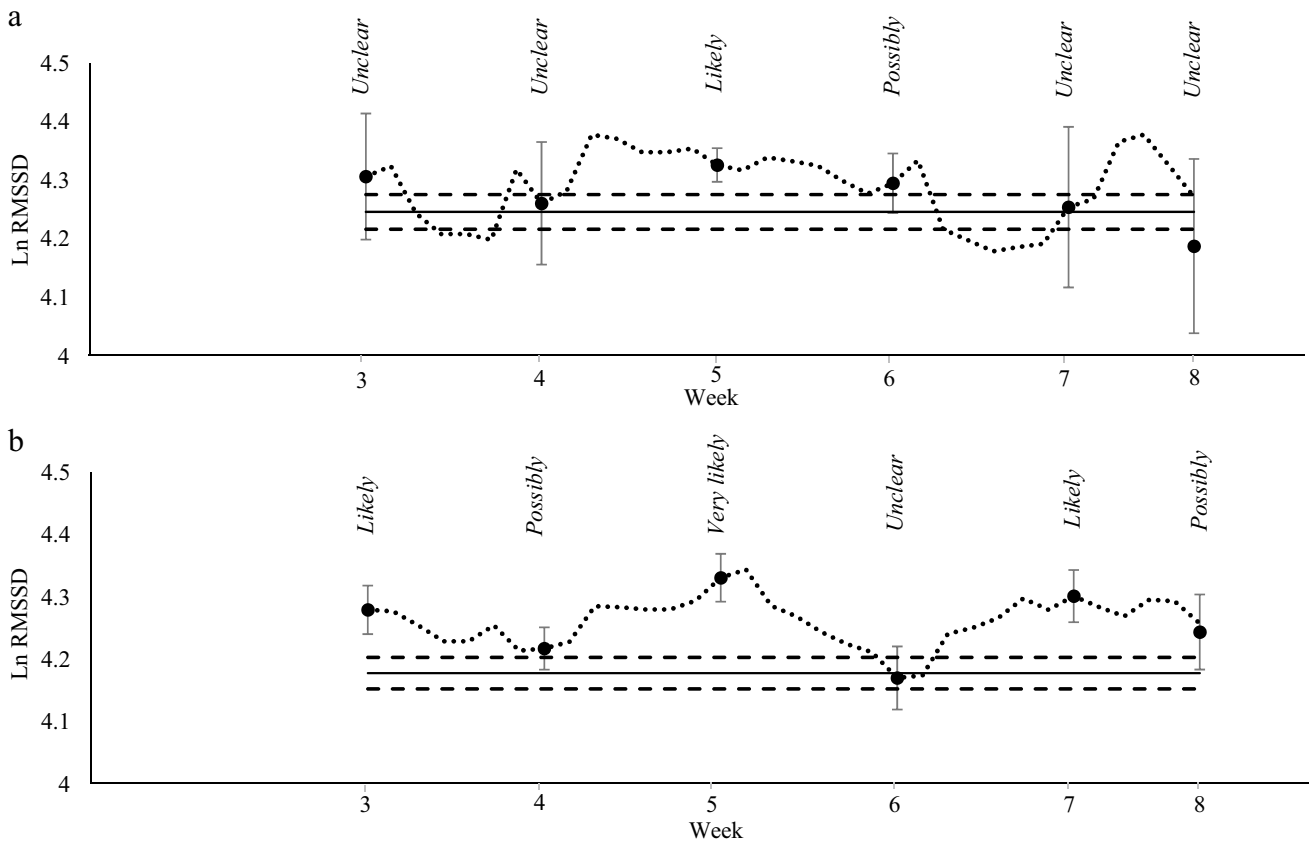
Fig. 3 Weekly Ln RMSSD changes of a random participant tracked with the Polar H10 chest strap synchronized with Kubios app (**a**) and with the Polar Nightly Recharge™ function present in the Polar Ignite sport watch (**b**). The solid line represents the zero line of the SWC to indicate unclear changes when 90% CI overlaps. The dashed lines represent the SWC (mean ± 0.2SD) [18]. The dotted line represents the seven days rolling average

Table 4 Magnitude-based weekly Ln RMSSD changes interpretation during an eight-week concurrent training program determined by two different devices (i.e. Kubios-H10 vs. Polar Ignite)

Ignite/H10	Unclear	Most likely ↓	Very likely ↓	Likely ↓	Possibly ↓	Possibly ↔	Likely ↔	Very likely ↔	Most likely ↔	Possibly ↑	Likely ↑	Very likely ↑	Most likely ↑	Total (weeks)
Unclear	5	1	1	4	1	2				2				16
Most likely ↓														0
Very likely ↓			1											1
Likely ↓	1		1	1	1					2		1		7
Possibly ↓	3			6		1				2	3	2		17
Possibly ↔				3										3
Likely ↔														0
Very likely ↔				1										1
Most likely ↔														0
Possibly ↑	4		1	1						2		1		9
Likely ↑	2				1					2		2	1	8
Very likely ↑				1						2			1	4
Most likely ↑														0
Total (weeks)	15	1	4	17	3	3	0	0	0	12	3	6	2	**66**

White diagonal squares: agreement (13.6%); dark grey squares: agreement in direction but discordance in the magnitude of the change (25.8%); light grey: discordance (60.6%) ↓: decrease; ↑: increase; ↔: unchanged

inter-beats intervals [7] and the level of agreement between PRV and HRV, obtained from pulse and R waves, has been tested in the early morning and during different sleep stages showing a high level of agreement [4, 25]. In particular, a recent study has determined the level of agreement between the Polar Nightly Recharge™ function and the Polar H10

chest strap in the Ln RMSSD monitoring during sleep in a group of recreationally-trained athletes [26]. Both devices recorded pulse and R waves intervals during the first four hours of sleep as established by default in this function of Polar sports watches [26]. A systematic bias and random error of 0.17 ± 0.40 ms were found being such differences interpreted by the authors as an appropriate accuracy for this PPG system [26]. Despite the different time frames applied in this study (i.e., night recording vs. upon awakening), our results reveal a systematic bias and random error of -0.05 ± 0.33 ms and -0.01 ± 0.28 ms for males and females, respectively, throughout the entire training program. This random error showed to greatly influence the Ln RMSSD weekly changes interpretation when the MBI was applied. Thus, from a practical point of view of their corresponding uses in different periods, there is a large discrepancy between devices.

One more important point that might explain the differences between systems is the signal processing. The software Kubios available in a phone application has shown to be a valid tool in this task by interpolating the identified artefacts using a cubic spline interpolation [8]. Benítez-Herrera et al. [27] compared the Kubios and Polar software in the inter-beat intervals correction artefacts during a maximum VO2max test in a group of 12 competitive athletes, finding a better performance in the Kubios software. However, since HRV measurement was conducted at rest, the artefacts correction should not be the main source of disparity between devices and software.

The MBI method applied in this manuscript was proposed by Batterham and Hopkins [10] in response to the limitation of the null-hypothesis significant testing (i.e., $p < 0.05$ despite the magnitude of the effect is trivial). However, its application in sport science is not without critics [28]. First, the same mean effect can have different magnitudes due to differences in the standard deviation [28] (e.g., Polar H10: 100 ± 20 ms to 120 ± 20 ms is ES $= 1.0$; Polar Ignite: 100 ± 40 ms to 120 ± 40 ms is ES $= 0.5$). However, in the case exposed, wider confident intervals reflect the particular variation of each system, and due to that, these differences need to be considered for determining the magnitude of the change. The random errors obtained from both devices through the training program (≈ 30 ms) could partially explain these differences when MBI is applied despite the large correlation ($r = 0.714$) and good reliability (ICC $= 0.817$) obtained. Second, establishing what is a worthwhile change can affect inferences [28]. In this case, the first two weeks of the training program were used to determine the SWC (mean ± 0.2SD) where the lightest micro-cycle load was applied. This baseline measure should be updated in longer training periods for meaningful HRV changes interpretation. Lastly, type I error rates are high in this method [28]. However, the quality assessment as

indicated in Table 1 warns about the probability of committing a wrong inferential [10]. These HRV change interpretations are with what coaches and athletes deal in daily training decisions, so the conducted analysis responds closely to the functionality of training monitoring with these devices.

Finally, some limitations must be taken into consideration to properly interpret these results. HRV daily monitoring was not conducted in a laboratory context and neither a gold standard was used. However, participants were instructed and encouraged to reduce any potential disturbance in the night and upon awaking measure. From a pragmatic approach, this situation was the real one where athletes and coaches would be exposed, so these potential artefacts on the measure should be considered when comparing wearable devices. It is also necessary to highlight the small number of participants recruited. However, the main results are derived from the 66 weeks monitored which provide some clear insights about their level of agreement. Based on this practical interpretation of the HRV modulation according to a baseline measure, future works should compare the level of agreement of other tracking procedures such as post-exercise monitoring and determine which one reflects more precisely the athlete state according to biological markers.

5 Conclusion

From a practical point of view, these results warn coaches and athletes about the influence of the method used to test and monitor HRV data and its interpretation as a measure of internal load. These advances in monitoring technologies are gaining in ecology, but there is a need to test their functionality before putting them into practice. Despite the very large correlation and good reliability obtained between devices, the random errors found conditioned the weekly Ln RMSSD changes interpretation when the MBI method was applied. Thus, practitioners in the field of training should be aware of these facts if their training groups use different devices or if an athlete interchanges them. The different nuances of each device (i.e., night recording vs. upon awakening measure, pulse vs. R waves, and software signal processing) can condition the HRV data variation what could compromise the interpretation of the ANS modulation.

References

1. Gabbett TJ, Nassis GP, Oetter E, Pretorius J, Johnston N, Medina D, Ryan A (2017) The athlete monitoring cycle: a practical guide to interpreting and applying training monitoring data. Br J Sports Med 51(20):1451–1452

2. Schäfer A, Vagedes J (2013) How accurate is pulse rate variability as an estimate of heart rate variability? A review on studies comparing photoplethysmographic technology with an electrocardiogram. Int J Cardiol 166(1):15–29

3. Buchheit M (2014) Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol 5:73

4. Plews DJ, Laursen PB, Stanley J, Kilding AE, Buchheit M (2013) Training adaptation and heart rate variability in elite endurance athletes: opening the door to effective monitoring. Sports Med 43(9):773–781

5. Polar (2022) Nightly recharge™ measure your recovery level. [online] Available at: https://support.polar.com/es/night. Accessed 26 Jan 2022

6. Selvaraj N, Jaryal A, Santhosh J, Deepak KK, Anand S (2008) Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography. J Med Eng Technol 32(6):479–484

7. Gilgen-Ammann R, Schweizer T, Wyss T (2019) RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. Eur J Appl Physiol 119(7):1525–1532

8. Tarvainen MP, Niskanen JP, Lipponen JA, Ranta-Aho PO, Karjalainen PA (2014) Kubios HRV–heart rate variability analysis software. Comput Methods Programs Biomed 113(1):210–220

9. Plews DJ, Laursen PB, Kilding AE, Buchheit M (2012) Heart rate variability in elite triathletes, is variation in variability the key to effective training? A case comparison. Eur J Appl Physiol 112(11):3729–3741

10. Batterham AM, Hopkins WG (2006) Making meaningful inferences about magnitudes. Int J Sports Physiol Perform 1(1):50–57

11. Plews DJ, Laursen PB, Le Meur Y, Hausswirth C, Kilding AE, Buchheit M (2014) Monitoring training with heart-rate variability: how much compliance is needed for valid assessment? Int J Sports Physiol Perform 9(5):783–790

12. Pescatello LS, Riebe D, Thompson PD (eds) (2014) ACSM's guidelines for exercise testing and prescription. Lippincott Williams & Wilkins

13. Hayashi M, Yoshikawa H, Uchiyama A, Higashino T (2019) Preliminary investigation on band tightness estimation of wrist-worn devices using inertial sensors. In: International Conference on Wireless Mobile Communication and Healthcare, pp 256–266

14. Malik M (1996) Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of cardiology and the North American Society of Pacing and Electrophysiology. Circulation 93:1043–1065

15. Hopkins W, Marshall S, Batterham A, Hanin J (2009) Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc 41(1):3

16. Bland JM, Altman D (1986) Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 327(8476):307–310

17. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficient for reliability research. J Chiropr Med 15(2):155–163

18. Cohen J (2013) Statistical power analysis for the behavioral sciences. Academic Press

19. Hopkins W (2021) Sportscience. [online] Sportsci.org. Available at: <http://www.sportsci.org. Accessed 15 Sep 2021

20. Hynynen ESA, Uusitalo A, Konttinen N, Rusko H (2006) Heart rate variability during night sleep and after awakening in overtrained athletes. Med Sci Sports Exerc 38(2):313

21. Myllymäki T, Kyröläinen H, Savolainen K, Hokka L, Jakonen R, Juuti T, Rusko H (2011) Effects of vigorous late-night exercise on sleep quality and cardiac autonomic activity. J Sleep Res 20(1pt2):146–153

22. Bent B, Goldstein BA, Kibbe WA, Dunn JP (2020) Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 3(1):1–9

23. Ruiz-Alias SA, García-Pinillos F, Soto-Hermoso VM, Ruiz-Malagón EJ (2021) Heart rate monitoring of the endurance runner during high intensity interval training: influence of device used on training functions. Proc Inst Mech Eng Part P: J Sports Eng Technol

24. Ruiz-Malagón EJ, Ruiz-Alias SA, García-Pinillos F, Delgado-García G, Soto-Hermoso VM (2020) Comparison between photoplethysmographic heart rate monitor from Polar Vantage M and Polar V800 with H10 chest strap while running on a treadmill: Validation of the Polar Precision Prime™ photoplestimographic system. Proc Inst Mech Eng Part P: J Sports Eng Technol 1754337120976659

25. Liu S, Teng J, Qi X, Wei S, Liu C (2017) Comparison between heart rate variability and pulse rate variability during different sleep stages for sleep apnea patients. Technol Health Care 25(3):435–445

26. Nuuttila OP, Korhonen E, Laukkanen J, Kyröläinen H (2022) Validity of the wrist-worn polar vantage V2 to measure heart rate and heart rate variability at rest. Sensors 22(1):137

27. Benítez-Herrera AB, Irurtia F, García-González MA (2017) Comparison of three software for heart rate variability artifact correction. Int Conf Sport Sci 1:1–1

28. Van Hooren B (2018) Magnitude-based inference: what is it? How does it work and is it appropriate? Sport Perf Sci 1:1–2